(a) Find the equation of the least-squares straight line through these points in the form $y = [m(\pm s_m)]x + [b(\pm s_b)]$ with a reasonable number of significant figures.

(b) Make a graph showing the experimental data and the calculated straight line.

(c) An unknown protein sample gave an absorbance of 0.973. Calculate the number of micrograms of protein in the unknown and estimate its uncertainty.

Problems

Gaussian Distribution

4-1. What is the relation between the standard deviation and the precision of a procedure? What is the relation between standard deviation and accuracy?

4-2. Use Table 4-1 to state what fraction of a Gaussian population lies within the following intervals:

(a) $\mu \pm \sigma$ **(c)** μ to $+\sigma$ **(e)** $-\sigma$ to -0.5σ
(b) $\mu \pm 2\sigma$ **(d)** μ to $+0.5\sigma$

4-3. The ratio of the number of atoms of the isotopes ^{69}Ga and ^{71}Ga in eight samples from different sources was measured in an effort to understand differences in reported values of the atomic mass of gallium:

Sample	^{69}Ga/^{71}Ga	Sample	^{69}Ga/^{71}Ga
1	1.526 60	5	1.528 94
2	1.529 74	6	1.528 04
3	1.525 92	7	1.526 85
4	1.527 31	8	1.527 93

SOURCE: *J. W. Gramlich and L. A. Machlan, Anal. Chem.* ***1985****, 57, 1788.*

Find the **(a)** mean, **(b)** standard deviation, and **(c)** variance. **(d)** Write the mean and standard deviation together with an appropriate number of significant digits.

4-4. (a) Calculate the fraction of bulbs in Figure 4-1 expected to have a lifetime greater than 1 005.3 h.

(b) What fraction of bulbs is expected to have a lifetime between 798.1 and 901.7 h?

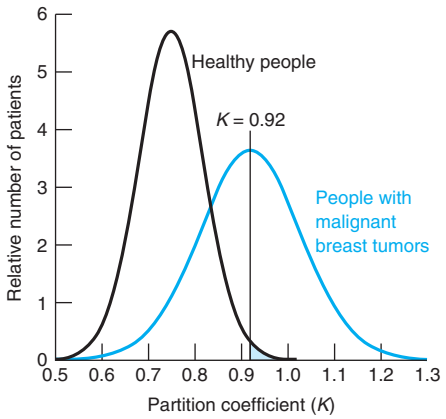
(c) Use the Excel NORMDIST function to find the fraction of bulbs expected to have a lifetime between 800 and 900 h.

4-5. Blood plasma proteins of patients with malignant breast tumors differ from proteins of healthy people in their solubility in the presence of various polymers. When the polymers dextran and poly(ethylene glycol) are mixed with water, a two-phase mixture is formed. When plasma proteins of tumor patients are added, the distribution of proteins between the two phases is different from that of plasma proteins of a healthy person. The distribution coefficient (K) for any substance is defined as K = [concentration of the substance in phase A] / [concentration of the substance in phase B]. Proteins of healthy people have a mean distribution coefficient of 0.75 with a standard deviation of 0.07. For the proteins of people with cancer, the mean is 0.92 with a standard deviation of 0.11.

(a) Suppose that K were used as a diagnostic tool and that a positive indication of cancer is taken as $K \geq 0.92$. What fraction of people with tumors would have a false negative indication of cancer because $K < 0.92$?

(b) What fraction of healthy people would have a false positive indication of cancer? This number is the fraction of healthy people with $K \geq 0.92$, shown by the shaded area in the adjoining graph. Estimate an answer with Table 4-1 and obtain a more exact result with the NORMDIST function in Excel.

(c) Vary the first argument of the NORMDIST function to select a distribution coefficient that would identify 75% of people with tumors. That is, 75% of patients with tumors would have K above the selected distribution coefficient. With this value of K, what fraction of healthy people would have a false positive result indicating they have a tumor?



Distribution coefficients of plasma proteins from healthy people and from people with malignant breast tumors. [Data from B. Y. Zaslavsky, "Bioanalytical Applications of Partitioning in Aqueous Polymer Two-Phase Systems," *Anal. Chem.* ***1992****, 64, 765A.*]

4-6. The equation for the Gaussian curve in Figure 4-1 is

$$y = \frac{(\text{total bulbs})(\text{hours per bar})}{s\sqrt{2\pi}} e^{-(x-\bar{x})^2/2s^2}$$

where $\bar{x}$ is the mean value (845.2 h), s is the standard deviation (94.2 h), total bulbs = 4 768, and hours per bar (= 20) is the width of each bar in Figure 4-1. Set up a spreadsheet like the one on the next page to calculate the coordinates of the Gaussian curve in Figure 4-1 from 500 to 1 200 h in 25-h intervals. Note the heavy use of parentheses in the formula at the bottom of the spreadsheet to force the computer to do the arithmetic as intended. Use Excel to graph your results.

4-7. Repeat Problem 4-6 but use the values 50, 100, and 150 for the standard deviation. Superimpose all three curves on a single graph.

Confidence Intervals, *t* Test, *F* Test, and Grubbs Test

4-8. What is the meaning of a confidence interval?

4-9. What fraction of vertical bars in Figure 4-5a is expected to include the population mean (10 000) if many experiments are carried out? Why are the 90% confidence interval bars longer than the 50% bars in Figure 4-5?

	A	B	C
1	Gaussian curve for light bulbs (Fig 4-1)		
2			
3	mean =	x (hours)	y (bulbs)
4	845.2	500	0.49
5	std dev =	525	1.25
6	94.2	550	2.98
7	total bulbs =	600	13.64
8	4768	700	123.11
9	hours per bar =	800	359.94
10	20	845.2	403.85
11	sqrt(2 pi) =	900	340.99
12	2.506628	1000	104.67
13		1100	10.41
14		1200	0.34
15	Formula for cell C4 =		
16	(A8*A10/(A6*A12))*		
17	EXP(-((B4-A4)^2)/(2*A6^2))		

Spreadsheet for Problem 4-6

4-10. List the three different cases that we studied for comparison of means, and write the equations used in each case.

4-11. The percentage of an additive in gasoline was measured six times with the following results: 0.13, 0.12, 0.16, 0.17, 0.20, 0.11%. Find the 90% and 99% confidence intervals for the percentage of the additive.

4-12. Sample 8 of Problem 4-3 was analyzed seven times, with $\bar{x} = 1.527\ 93$ and $s = 0.000\ 07$. Find the 99% confidence interval for sample 8.

4-13. A trainee in a medical lab will be released to work on her own when her results agree with those of an experienced worker at the 95% confidence level. Results for a blood urea nitrogen analysis are shown below.

Trainee: $\bar{x} = 14.5_7$ mg/dL $s = 0.5_3$ mg/dL $n = 6$ samples

Experienced worker: $\bar{x} = 13.9_5$ mg/dL $s = 0.4_2$ mg/dL $n = 5$ samples

(a) What does the abbreviation dL stand for?

(b) Should the trainee be released to work alone?

4-14. The CdSe content (g/L) of nanocrystals was measured by two methods for six different samples. Do the two methods differ significantly at the 95% confidence level?

Sample	Method 1 Anodic stripping	Method 2 Atomic absorption
A	0.88	0.83
B	1.15	1.04
C	1.22	1.39
D	0.93	0.91
E	1.17	1.08
F	1.51	1.31

SOURCE: *E. Kuçur, F. M. Boldt, S. Cavaliere-Jaricont, J. Ziegler, and T. Nann, Anal. Chem.* ***2007****, 79, 8987.*

4-15. Now we use a built-in routine in Excel for the paired *t* test to see if the two methods in Problem 4-14 produce significantly different results. Enter the data for Methods 1 and 2 into two columns of a spreadsheet. For Excel 2007, find Data Analysis in the Data ribbon. In earlier versions of Excel, find Data Analysis in the Tools menu. If Data Analysis does not appear, follow the instructions at the beginning of Section 4-5 to load this software. Select Data Analysis and then select t-Test: Paired Two Sample for Means. Follow the instructions of Section 4-5, and the routine will print out information including $t_{calculated}$ (labeled t Stat) and t_{table} (labeled t Critical two-tail). You should reproduce the results of Problem 4-14.

4-16. Two methods were used to measure fluorescence lifetime of a dye. Are the standard deviations significantly different? Are the means significantly different?

Quantity	Method 1	Method 2
Mean lifetime (ns)	1.382	1.346
Standard deviation (ns)	0.025	0.039
Number of measurements	4	4

SOURCE: *N. Boens et al., Anal. Chem.* ***2007****, 79, 2137.*

4-17. Do the following two sets of measurements of $^6Li/^7Li$ in a Standard Reference Material give statistically equivalent results?

Method 1	Method 2
0.082 601	0.081 83
0.082 621	0.081 86
0.082 589	0.082 05
0.082 617	0.082 06
0.082 598	0.082 15
	0.082 08

SOURCE: *S. Ahmed, N. Jabeen, and E. ur Rehman, Anal. Chem.* ***2002****, 74, 4133; L. W. Green, J. J. Leppinen, and N. L. Elliot, Anal. Chem.* ***1988****, 60, 34.*

4-18. If you measure a quantity four times and the standard deviation is 1.0% of the average, can you be 90% confident that the true value is within 1.2% of the measured average?

4-19. Students measured the concentration of HCl in a solution by titrating with different indicators to find the end point.

Indicator	Mean HCl concentration (M) (± standard deviation)	Number of measurements
1. Bromothymol blue	0.095 65 ± 0.002 25	28
2. Methyl red	0.086 86 ± 0.000 98	18
3. Bromocresol green	0.086 41 ± 0.001 13	29

SOURCE: *D. T. Harvey, J. Chem. Ed.* ***1991****, 68, 329.*

Is the difference between indicators 1 and 2 significant at the 95% confidence level? Answer the same question for indicators 2 and 3.

4-20. Hydrocarbons in the cab of an automobile were measured during trips on the New Jersey Turnpike and trips through the Lincoln Tunnel connecting New York and New Jersey.[10] The concentrations (± standard deviations) of *m*- and *p*-xylene were

Turnpike: $31.4 \pm 30.0\ \mu g/m^3$ (32 measurements)

Tunnel: $52.9 \pm 29.8\ \mu g/m^3$ (32 measurements)

Do these results differ at the 95% confidence level? At the 99% confidence level?

4-21. A Standard Reference Material is certified to contain 94.6 ppm of an organic contaminant in soil. Your analysis gives values of 98.6, 98.4, 97.2, 94.6, and 96.2 ppm. Do your results differ from the expected result at the 95% confidence level? If you made one more measurement and found 94.5, would your conclusion change?

4-22. Nitrite (NO_2^-) was measured by two methods in rainwater and unchlorinated drinking water. The results ± standard deviation (number of samples) are

Sample source	Gas chromatography	Spectrophotometry
Rainwater	0.069 ± 0.005 mg/L (n = 7)	0.063 ± 0.008 mg/L (n = 5)
Drinking water	0.078 ± 0.007 mg/L (n = 5)	0.087 ± 0.008 mg/L (n = 5)

SOURCE: *I. Sarudi and I. Nagy, Talanta* ***1995***, *42, 1099.*

(a) Do the two methods agree with each other at the 95% confidence level for both rainwater and drinking water?

(b) For each method, does the drinking water contain significantly more nitrite than the rainwater (at the 95% confidence level)?

4-23. Should the value 216 be rejected from the set of results 192, 216, 202, 195, and 204?

Linear Least Squares

4-24. A straight line is drawn through the points (3.0, -3.87×10^4), (10.0, -12.99×10^4), (20.0, -25.93×10^4), (30.0, -38.89×10^4), and (40.0, -51.96×10^4) to give $m = -1.298\ 72 \times 10^4$, $b = 256.695$, $s_m = 13.190$, $s_b = 323.57$, and $s_y = 392.9$. Express the slope and intercept and their uncertainties with reasonable significant figures.

4-25. Here is a least-squares problem that you can do by hand with a calculator. Find the slope and intercept and their standard deviations for the straight line drawn through the points $(x,y) = (0,1)$, (2,2), and (3,3). Make a graph showing the three points and the line. Place error bars ($\pm s_y$) on the points.

4-26. Set up a spreadsheet to reproduce Figure 4-15. *Add error bars:* Follow the procedure on page 90. Use s_y for the + and – error.

4-27. *Excel LINEST function.* Enter the following data in a spreadsheet and use LINEST to find slope, intercept, and standard deviations. Use Excel to draw a graph of the data and add a trendline. Draw error bars of $\pm s_y$ on the points.

x:	3.0	10.0	20.0	30.0	40.0
y:	–0.074	–1.411	–2.584	–3.750	–5.407

Calibration Curves

4-28. Explain the following statement: "The validity of a chemical analysis ultimately depends on measuring the response of the analytical procedure to known standards."

4-29. Suppose that you carry out an analytical procedure to generate a linear calibration curve like that shown in Figure 4-13. Then you analyze an unknown and find an absorbance that gives a negative concentration for the analyte. What does this mean?

4-30. Using the linear calibration curve in Figure 4-13, find the quantity of unknown protein that gives a measured absorbance of 0.264 when a blank has an absorbance of 0.095.

4-31. Consider the least-squares problem in Figure 4-11.

(a) Suppose that a single new measurement produces a y value of 2.58. Find the corresponding x value and its uncertainty.

(b) Suppose you measure y four times and the average is 2.58. Calculate the uncertainty based on four measurements, not one.

4-32. Consider the linear calibration curve in Figure 4-13, which is derived from the 14 corrected absorbances in the shaded region at the right side of Table 4-7. Create a least-squares spreadsheet like Figure 4-15 to compute the equation of the line and the standard deviations of the parameters. Suppose that you find absorbance values of 0.265, 0.269, 0.272, and 0.258 for four identical samples of unknown and absorbances of 0.099, 0.091, 0.101, and 0.097 for four blanks. Find the corrected absorbance by subtracting the average blank from the average absorbance of the unknown. Calculate the amount of protein and its uncertainty in the unknown.

4-33. Here are mass spectrometric signals for methane in H_2:

CH_4 (vol%):	0	0.062	0.122	0.245	0.486	0.971	1.921
Signal (mV):	9.1	47.5	95.6	193.8	387.5	812.5	1 671.9

(a) Subtract the blank value (9.1) from all other values. Then use the method of least squares to find the slope and intercept and their uncertainties. Construct a calibration curve.

(b) Replicate measurements of an unknown gave 152.1, 154.9, 153.9, and 155.1 mV, and a blank gave 8.2, 9.4, 10.6, and 7.8 mV. Subtract the average blank from the average unknown to find the average corrected signal for the unknown.

(c) Find the concentration of the unknown and its uncertainty.

4-34. *Nonlinear calibration curve.* Following the procedure in Box 4-2, find how many micrograms (μg) of protein are contained in a sample with a corrected absorbance of 0.350 in Figure 4-13.

4-35. *Logarithmic calibration curve.* Calibration data spanning five orders of magnitude for an electrochemical determination of *p*-nitrophenol are given in the table. (The blank has already been subtracted from the measured current.) If you try to plot these data on a linear graph extending from 0 to 310 μg/mL and from 0 to 5 260 nA, most of the points will be bunched up near the origin. To handle data with such a large range, a logarithmic plot is helpful.

p-Nitrophenol (μg/mL)	Current (nA)	*p*-Nitrophenol (μg/mL)	Current (nA)
0.010 0	0.215	3.00	66.7
0.029 9	0.846	10.4	224
0.117	2.65	31.2	621
0.311	7.41	107	2 020
1.02	20.8	310	5 260

SOURCE: *Data from Figure 4 of L. R. Taylor, Am. Lab., February 1993, p. 44.*

(a) Make a graph of log(current) versus log(concentration). Over what range is the log-log calibration linear?

(b) Find the equation of the line in the form log (current) = $m \times$ log (concentration) + b.

(c) Find the concentration of *p*-nitrophenol corresponding to a signal of 99.9 nA.

4-36. *Confidence interval for calibration curve.* To use a calibration curve based on n points, we measure a new value of y and calculate the corresponding value of x. The one-standard-deviation uncertainty in x, s_x, is given by Equation 4-27. We express a *confidence interval* for x, using Student's t:

$$\text{Confidence interval} = x \pm ts_x$$

where t is taken from Table 4-2 for $n - 2$ degrees of freedom.

A calibration curve based on $n = 10$ known points was used to measure the protein in an unknown. The results were protein = 15.2_2 ($\pm 0.4_6$) μg, where $s_x = 0.4_6$ μg. Find the 90% and 99% confidence intervals for protein in the unknown.